Machine Learning 1

Lecture 12.1 - Kernel Methods
Gaussian Processes - Properties of Gaussian Random Variables

Erik Bekkers

(Bishop 2.3.1, 2.3.2)

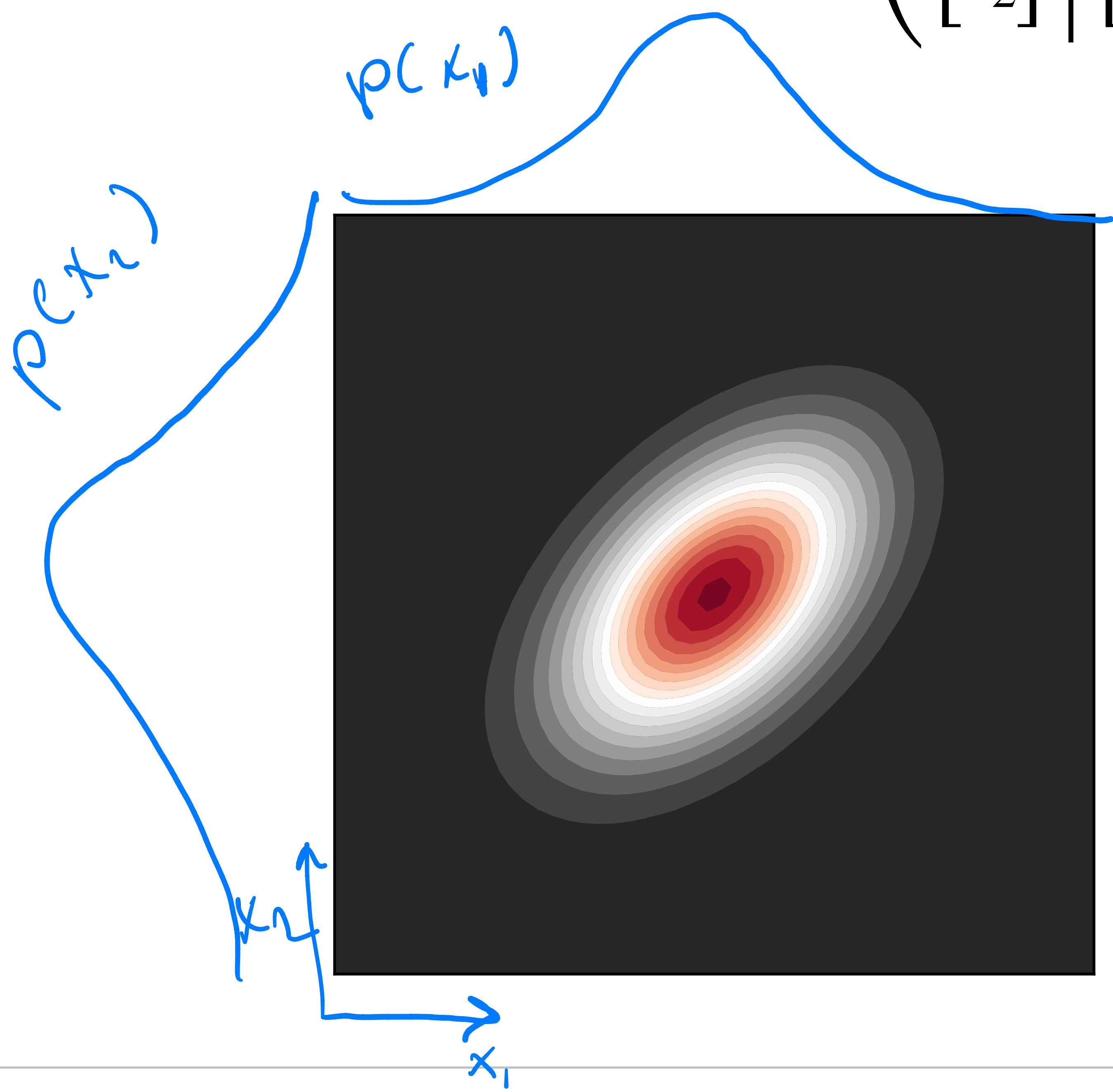
Slide credits: Rianne van den Berg

Image credit: Kirillm | Getty Images

Gaussians: Marginalization property

- Take two random variables x_1 and x_2, that are jointly Gaussian distributed:

$$p(x_1, x_2) = \mathcal{N}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \middle| \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right)$$



- Then the marginals are given by

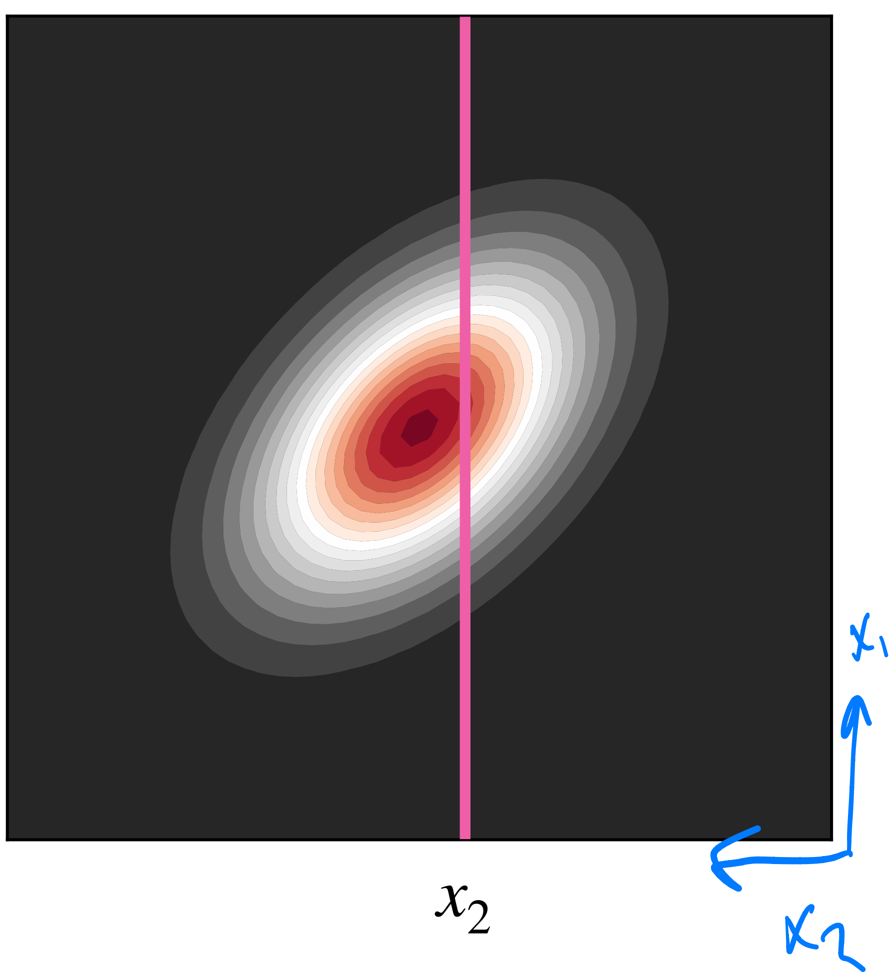
$$p(x_1) = \mathcal{N}(x_1 \mid \mu_1, \Sigma_{11})$$

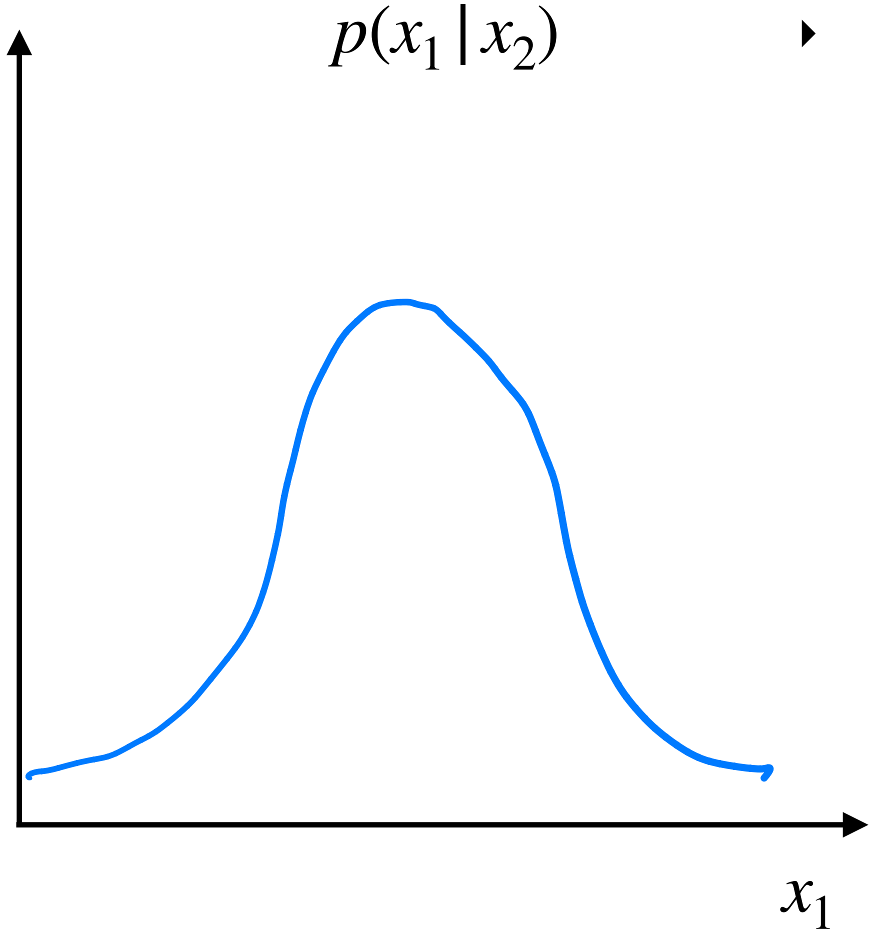
$$p(x_2) = \mathcal{N}(x_2 \mid \mu_2, \Sigma_{22})$$

Gaussians: Conditioning Property

- Take two random variables x_1 and x_2, that are jointly Gaussian distributed:

$$p(x_1, x_2) = \mathcal{N}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \middle| \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right)$$





- Then the conditional is:

$$p(x_1 \mid x_2) = \mathcal{N}(\mu_{1|2}, \Sigma_{1|2})$$

with

$$\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$$

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

Summing Random Variables

- The sum of two independent Gaussian random variables is also a Gaussian random variable:
  - If

$$x \sim \mathcal{N}(\mu, \Sigma)$$

$$y \sim \mathcal{N}(\mu', \Sigma')$$

x, y independent!

  - Then

$$z = x + y \qquad \rightarrow \qquad z \sim \mathcal{N}(\mu + \mu', \Sigma + \Sigma')$$

Sampling Correlated Gaussian Variables

- If we have sampled a vector $\mathbf{x}$ of uncorrelated Gaussian variables:
- $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

Reparametrization trick

- And if $\mathbf{y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{x}$
- Then $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \underbrace{\mathbf{A}\mathbf{A}^T}_{\Sigma})$ with $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T$
- So if you have access to a sampler for uncorrelated Gaussian variables, you can create correlated samples for a given mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$

$y \sim N(\mu, \Sigma)$

- *For a given $\boldsymbol{\Sigma}$, you can compute $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T$ with a Cholesky decomposition such that $\mathbf{A}$ is lower triangular*
- *Or you compute the eigendecomposition $\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$ and take $\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}^{1/2}$*

Sampling Correlated Gaussian Variables

- If $\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}$ and $p(\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}) = \mathcal{N}\left(\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \middle| \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right)$

- Then the marginals are given by

$$p(\mathbf{x}_1) = \mathcal{N}\left(\mathbf{x}_1 \,|\, \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}\right)$$

$$p(\mathbf{x}_2) = \mathcal{N}\left(\mathbf{x}_2 \,|\, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}\right)$$

- And the conditional is given by

$$p(\mathbf{x}_1 \,|\, \mathbf{x}_2) = \mathcal{N}\left(\mathbf{x}_1 \,|\, \boldsymbol{\mu}_{1|2}, \boldsymbol{\Sigma}_{1|2}\right) \quad \text{with} \quad \boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2)$$

$$\boldsymbol{\Sigma}_{1|2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$$

- If $\mathbf{x}$ is an uncorrelated Gaussian random variable (i.e., $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$) then $\mathbf{y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{x}$ is correlated $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{A}\mathbf{A}^T)$ with $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T$